Osteoarthritis and Cartilage 19 (2011) 324–331

Osteoarthritis and Cartilage

Histopathology grading systems for characterisation of human knee osteoarthritis – reproducibility, variability, reliability, correlation, and validity

R.G. Pearson*, T. Kurien, K.S.S. Shu, B.E. Scammell

University of Nottingham, Division of Orthopaedic & Accident Surgery, School of Clinical Sciences, Queen's Medical Centre Campus, Nottingham NG7 2UH, UK

ARTICLE INFO

Article history:
Received 6 May 2010
Accepted 6 December 2010

Keywords:
Osteoarthritis
Cartilage
Grading
Histopathology
Reproducibility
Reliability

SUMMARY

Objective: To determine the reliability, reproducibility, variability and validity of the Osteoarthritis Cartilage Histopathology (OACH) assessment system and Mankin Histological–Histochemical Grading System (HHGS) when applied to the characterisation of the osteoarthritic human knee.
Method: Osteoarthritic knees of 10 patients undergoing unilateral knee arthroplasty were assessed, and assigned Kellgren–Lawrence and Line Drawing Atlas (LDA) radiology scores. The tibial plateaux were scored using the Modified Collins (MC) and Société Française d'Arthroscopie (SFA). Three observers twice scored both the OACH and HHGS across a single complete medial and lateral tibial plateau transect taken to include the region with the most severe osteoarthritis (OA) lesion. Intra- and inter-observer reliability, reproducibility, variability and validity were assessed, and the correlation between the two histopathology scoring systems was calculated.
Result: Both histopathology scoring systems were determined to be reliable and reproducible exhibiting similar variability, when applied to characterise OA specimens sampled from a well defined patient group with knee OA. A strong correlation between the mean OACH and mean HHGS scores was identified (Spearman's ρ 0.980, $P < 0.0001$).
Conclusion: Both scoring systems implemented provide useful measures in the characterisation of knee osteoarthritis. It is of note that an additional parameter within the OACH score over the HHGS defines the extent of the disease, where the HHGS is a grade attributed to the most representative level of the biological aggression within the OA lesions. This study has confirmed the OACH system's utility in human knee OA and is supported by a significant correlation with the established HHGS.

Introduction

A comprehensive and reproducible histopathology grading scale with universal acceptance would be of significant benefit as a research tool for both comparison of osteoarthritis (OA) histopathology data, and reference to other OA modes of imaging such as X-ray or magnetic resonance imaging (MRI)[1]. The OA Histological Histochemical Grading System (HHGS) introduced by Mankin *et al.* in 1971 is currently the most commonly applied scale[2]. This system was developed to describe human hip OA but has subsequently been applied directly or with modification to other synovial joints and a range of animal models.

The reproducibility of the HHGS scoring system has been described as inadequate and there is inconsistent opinion regarding the reliability[3–5]. The ability to discriminate between normal and mild to moderate OA histopathology by the HHGS scoring system has also been questioned, and the use of end-stage diseased specimens to establish the system has been proposed to account for the lack of sensitivity and specificity[3,6]. A fundamental problem when validating an OA histopathology scoring system is the identification of robust criteria to provide definitive values for reference when defining the severity of the osteoarthritic lesion. Cartilage of cadaveric origin described as macroscopically normal and end-stage OA tibial plateaux from arthroplasty of Collins Grades III and IV have been used as validation criteria in HHGS validation experiments[3,6].

The Osteoarthritis Research Society International (OARSI) through the establishment in 1998 of an OA Working Group introduced the Osteoarthritis Cartilage Histopathology (OACH) scoring system with the intention of improving established systems based on contemporary pathophysiological knowledge[7]. The reliability, reproducibility and variability of the system has been tested and compared to the HHGS in an animal model where OA was surgically induced by a cobalt–chrome implant placed in the medial tibial plateau of the right knee of four Dutch milk goats. The

* Address correspondence and reprint requests to: Richard G. Pearson, University of Nottingham, Division of Orthopaedic & Accident Surgery, School of Clinical Sciences, Queen's Medical Centre Campus, Nottingham NG7 2UH, UK. Tel: 44-115-8231119; Fax: 44-115-8231118.
E-mail address: richard.pearson@nottingham.ac.uk (R.G. Pearson).

1063-4584/$ – see front matter
doi:10.1016/j.joca.2010.12.005

authors concluded that the reliability of the OACH was greater than the HHGS[5].

To establish the utility of the OACH for assessing OA in humans we examined 10 osteoarthritic tibial plateaux obtained at total knee arthroplasty (TKA). Statistical analysis defined reproducibility, variability and reliability. The same specimens were also scored using the HHGS to enable comparison. The validity of the histopathology grading systems was also examined using the Société Française d'Arthroscopie (SFA) chondropathy scoring system, and the radiological Kellgren–Lawrence (K–L) and Line Drawing Atlas (LDA) scores.

Methods

Patient details

Tibial plateaux were obtained from 10 OA patients undergoing TKA at the Nottingham University Hospitals. The group was comprised of six females with a mean age 73 (range 51–79) and four males with mean age 63 (55–71). This research was conducted according to the 1975 (revised 2008) World Medical Association Declaration of Helsinki guidelines and was approved by the Nottingham Research Ethics Committee (approval number 06/Q2404/163). All patients gave written informed consent. Oxford knee pain scores (revised scale) ranged from 16 to 38 with a mean value of 30.6, using the 0–48 range where 48 indicates the patient perceiving the least pain[8]. The synovium was visually graded (0–3, normal, mild, moderate and severe) for inflammation during TKA surgery and all synovia were either moderately or severely inflamed. X-rays identified chondrocalcinosis of the knee in two patients.

Clinical characterisation

Clinical severity of knee OA was assessed by X-ray, using K–L and LDA grades[9,10]. The patients' drugs were recorded and Oxford knee scores were derived from the self-administered questionnaire to provide an index of the patient-perceived pain. Tibial plateaux at TKA were placed in 10% neutral buffered formalin. The extent of OA involvement was characterised using the Modified Collins (MC)[11,12] and SFA[13], these visual scoring systems were applied prior to processing the tibial plateaux for histopathology scoring (Fig. 1). The extent of each SFA grade in each tibial plateau compartment was recorded on a diagram of the tibial plateau and an estimate of the percentage of the articular surface occupied by each grade was entered into a table (Fig. 1), Grade 0 = normal, Grade I = swelling and softening, Grade II = superficial fibrillation, Grade III = deep fibrillation and Grade IV = exposure of subchondral bone. These values were input into the published SFA formulae for medial and lateral compartments respectively[13].

Histology processing

Post visual chondropathy grading of the OA, the tibial plateaux were processed to enable histopathological evaluation of OA of the articular surface using the OACH and HHGS scoring systems. A photographic archive of the tibial plateaux facilitated correct specimen orientation throughout histopathology processing. Coronal osteochondral specimen blocks (3 mm in diameter) were cut using a diamond impregnated rotary slitting saw (Material Science Medical, UK) through the entire tibial plateau from the medial to lateral compartment periphery, sequentially from the anterior to the posterior of each plateau (Fig. 1). Osteochondral specimens were fixed for 48 h in >10 volumes 10% neutral buffered formalin. Decalcification was conducted for 48 h using a Sakura

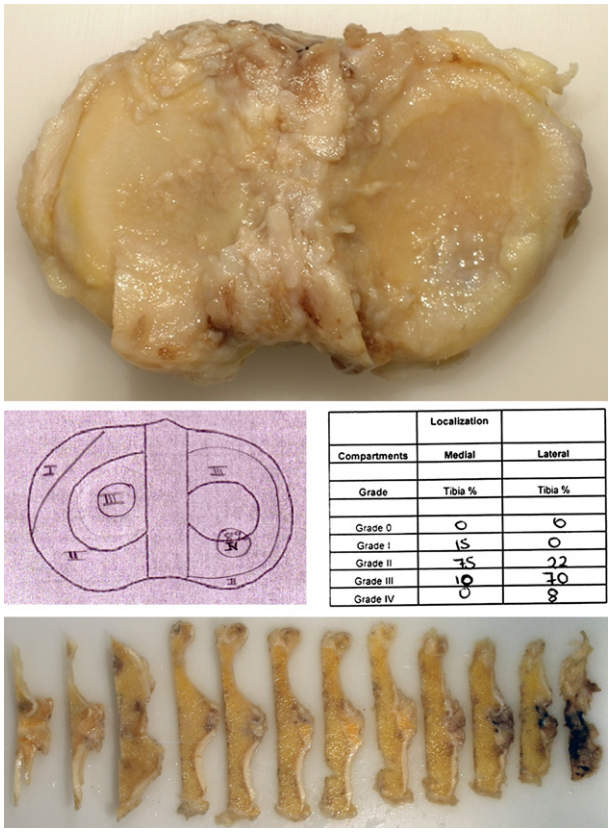

Fig. 1. Macroscopic chondropathy on a tibial plateau articular surface (top), SFA scoring of OA grades on schematic diagram within the patient information proforma and the table containing attributed percentage surface coverage of grades to the lateral and medial tibial compartments (centre); osteochondral specimens (3 mm blocks) generating a sequential coronal anterior–posterior series (bottom).

proprietary decalcification solution within a Sakura TDE decalcifier system (Sakura Finetek Europe BV, NL). Specimens were dehydrated using a series of graded alcohols, cleared in xylene and embedded in histological grade paraffin wax. 5 μm sections were stained with haematoxylin and eosin, and serial sections were stained with Safranin O and Fast Green[14].

Histopathology scoring systems

The osteochondral specimen that transected the centre of the most severe OA lesion in each tibial plateau was scored. Both the medial and lateral compartments were scored. This was for two main reasons (1), the predominant OA lesion was not always on the medial condyle and (2), the contralateral tibial compartment to the most severe OA lesion provided specimens of less severe OA, often with mild osteoarthritic histopathology suiting validation that requires a complete range of OA severity.

All observers examined OA osteochondral histopathology specimens originating from a similar patient group to those in the study. This was in the format of two half day sessions where details of applying HHGS and OACH were discussed. Data for the OACH and HHGS scoring system was compiled from three observers (RP, BES, TK), rating each microscope slide ($n = 60$) on two occasions with an interval of between 2 and 3 weeks between rating the specimens with the same scoring system and 1 or 2 weeks between ratings of different scoring systems. The scoring schemes are outlined for both grade and stage of OACH in Tables I and II, and the HHGS in Table III[2,7]. The OACH score is the product of the grade and the stage

Table I
OA cartilage histopathology grade[2]

Grade 0: surface intact, cartilage morphology intact	Matrix: normal architecture Cells: intact, appropriate orientation
Grade 1: surface intact	Matrix: superficial zone intact, oedema and/or superficial fibrillation (abrasion), focal superficial matrix condensation Cells: death, proliferation (clusters), hypertrophy, superficial zone Reaction must be more than superficial fibrillation only
Grade 2: surface discontinuity	As Grade 1 + Matrix discontinuity at superficial zone (deep fibrillation) ± Cationic stain matrix depletion (Safranin O or Toluidine Blue) upper 1/3 of cartilage ± Focal perichondronal increased stain (mid zone) ± Disorientation of chondron columns Cells: death, proliferation (clusters), hypertrophy
Grade 3: vertical fissures (clefts)	As Grade 1 Matrix vertical fissures into mid zone, branched fissures ± Cationic stain depletion (Safranin O or Toluidine Blue) into lower 2/3 of cartilage (deep zone) ± New collagen formation (polarized light microscopy, Picro Sirius Red stain) Cells: death, regeneration (clusters), hypertrophy, cartilage domains adjacent to fissures
Grade 4: erosion	Cartilage matrix loss: delamination of superficial layer, mid layer cyst formation Excavation: matrix loss superficial layer and mid zone
Grade 5: denudation	Surface: sclerotic bone or reparative tissue including fibrocartilage within denuded surface Microfracture with repair limited to bone surface
Grade 6: deformation	Bone remodelling (more than osteophyte formation only). Includes: microfracture with fibrocartilaginous and osseous repair extending above the previous surface

OACH score is the product of the grade and the stage (cf. Table II).

and therefore the scale ranges from 0 to 24 with the HHGS scale range between 0–14. The OACH scoring system applied the grade and stage, and therefore incorporated a factor for the surface extent of the most severe lesion within each of the medial and lateral compartments, whereas the HHGS assessed the most representative region, as determined by each observer independently, within the medial and lateral compartments. The most representative region being defined as the grade that extended over the largest percentage of the specimen transect. A calibrated reticule was available when required to distinguish between two similar sized lesions of different grades. If the histology specimen was denuded of cartilage exhibiting eburnated bone the HHGS assigned a value of 14 if this was the most representative histopathology of the compartment transect.

Statistical analysis

Agreement between first and second scoring occasions was illustrated in a similar manner to Bland and Altman plots, where individual scores were plotted as values plus or minus from the mean for each compartment and for all observers[15]. 95% limits of agreement were plotted, ±1.96 standard deviations from the mean. Funnel or cone effects in these plots are apparent when there are greater or lesser discrepancies form the mean value at high or low values on the OA histopathology scale. Differences between the first and second occasion the specimens were scored investigated bias between scoring occasions. The intra-observer and inter-observer reproducibility was investigated using the intraclass correlation coefficient (ICC). This was calculated using a two-way random effects analysis of variance (ANOVA), type consistency, single measures[16].

Table II
OACH osteoarthritic cartilage histopathology stage assessment[2]

	% Involvement (surface, area, volume)
Stage 0	No OA activity seen
Stage 1	<10%
Stage 2	10–25%
Stage 3	25–50%
Stage 4	>50%

OACH score is the product of the grade (cf. Table I) and the stage.

Reliability of the data was determined using the average measures ICC, two-way random effects ANOVA, type consistency, the same calculation that is performed for determining Cronbach's alpha when investigating internal consistency of a measurement scale such as a questionnaire[17]. Correlation between the OACH and HHGS was tested using the Spearman's ρ statistic. Statistical analysis used SPSS 14.0 and GraphPad Prism 5.02 (agreement) software.

Results

Plots assessing intra-observer agreement between the first and second scores for either HHGS or OACH did not identify funnel or cone effects, therefore the data did not require transformation prior

Table III
Histology histopathology grading system[7]

	Grade
I. Structure	
a. Normal	0
b. Surface irregularities	1
c. Pannus and surface irregularities	2
d. Clefts to transitional zone	3
e. Clefts to radial zone	4
f. Clefts to calcified zone	5
g. Complete disorganisation	6
II. Cells	
a. Normal	0
b. Diffuse hypercellularity	1
c. Cloning	2
d. Hypocellularity	3
III. Safranin-O staining	
a. Normal	0
b. Slight reduction	1
c. Moderate reduction	2
d. Severe reduction	3
e. No dye noted	4
IV. Tidemark integrity	
a. Intact	0
b. Crossed by blood vessels	1

Mankin HHGS score is the sum of structure, cells, Safranin-O staining and tidemark integrity.

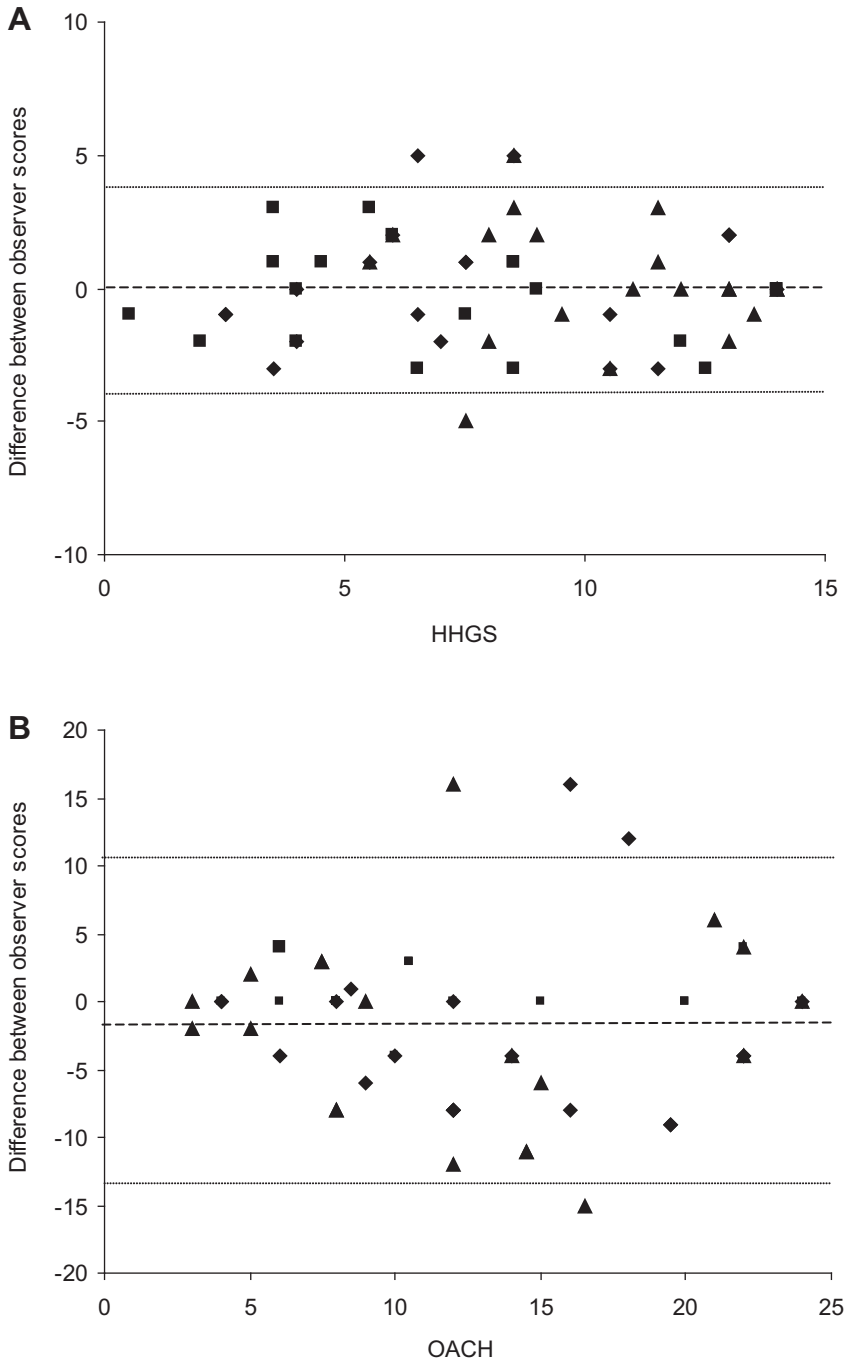

Fig. 2. HHGS and OACH system limits of agreement. Plots do not exhibit funnel or cone effects, the differences between first and second scores were plotted against the mean of these values as determined for each observer. These differences were consistent through the range of HHGS A and OACH B values. Observer A (◆), B (■), and C (▲); fine dotted lines represent overall 95% limits of agreement.

to analysis (Fig. 2). Bias between first and second scoring events was minimal with overall intra-observer bias of 0.07 and −1.32, HHGS and OACH respectively. Intra-observer agreement for HHGS was similar for the three observers, with 95% limits of agreement of (−4.5, 4.9), (−3.9, 3.4) and (−4.2, 4.7), and variable (−15.3, 10.6), (−3.1, 4.8) and (−16.9, 12.0) for the OACH (Table IV).

Intra- and inter-observer reproducibility was assessed using the ICC, single measures. The mean intra-observer reproducibility between the first and second scoring for both the HHGS and OACH, was 0.803 and 0.668 respectively. Similarly, the HHGS ICC for inter-observer reproducibility was 0.775 and 0.730 for first and second ratings. However the inter-observer OACH values indicated more variability in both first and second ratings, 0.448 and 0.409 respectively (Table V). Observers A and B were both experienced in the application of the HHGS. No rater was experienced in the attribution of the OACH prior to this research; and this may have contributed to the OACH inter-observer variation.

Validation of a histopathology scoring system is complicated by requiring a definitive measure of osteoarthritic pathology as a baseline. When HHGS vs OACH as mean values of the two scoring occasions for each observer was plotted there was a marked spread in the data [Fig. 3(A)]. These data were complemented by

Table IV
Limits of agreement between scoring occasions

Intra-observer agreement			
	Δ1 (bias)	SD	95% limits of agreement
HHGS			
Observer A	0.20	2.38	−4.46 4.86
Observer B	−0.25	1.86	−3.90 3.40
Observer C	0.25	2.29	−4.24 4.74
Overall	0.07	2.16	−4.17 4.30
OACH			
Observer A	−2.35	6.59	−15.26 10.56
Observer B	0.85	2.01	−3.08 4.78
Observer C	−2.45	7.35	−16.85 11.95
Overall	−1.32	5.92	−12.92 10.28

Δ1 Mean difference between the first and second scores.

Spearman's ρ statistics of 0.775, 0.806 and 0.906 for the individual observers A, B & C respectively, and this correlation was sustained in the overall correlation between the HHGS and the OACH, 0.757, $P < 0.0001$. In contrast, when the overall mean values for HHGS and OACH were plotted a linear relationship was observed with a regression coefficient of 0.922 fitting the data, Spearman's ρ 0.980, $P < 0.0001$ [Fig. 3(B)].

Histopathology validation was also provided by comparison with the SFA visual chondropathy scoring system. The SFA chondropathy scores for the medial compartment gave a mean of 244 with 95% confidence interval (CI) (214, 274), and a lateral compartment mean 270 with 95% CI (200, 341). A correlation was identified between SFA score and OACH histopathology score, Spearman's ρ 0.601, $P = 0.005$ [Fig. 3(C)]. K–L scores identified the degree of involvement of the different knee compartments with the medial compartment being most consistently involved with a mean score of 3.1 (range 3–4), and lateral 2.5 (1–4). Radiological scores did not include the patella (K–L) or patellofemoral joint space or patellofemoral osteophyte scores from flexed skyline X-ray view (LDA). There was correlation between the SFA chondropathy score and the radiological scores. For the SFA and K–L, Spearman's ρ 0.537, $P = 0.015$ [Fig. 3(D)], with SFA and LDA radiological scores having a similar correlation (Spearman's ρ 0.573, $P = 0.008$, data not shown).

Discussion

This research assessed the utility of the OACH when applied to human knee OA, and is the first report validating this histopathology scoring system using human specimens[5]. In this process we characterised osteochondral specimens from human osteoarthritic tibial plateaux using both the HHGS and OACH scoring systems. We determined that both these OA scoring systems had similar reproducibility, variability, and reliability, based on data from three independent observers assessing each specimen twice using each of the histopathology assessment systems. Essentially, the HHGS examines specific criteria independently; structure, chondrocyte status, Safranin-O staining of glycosaminoglycans, and the integrity of the tidemark, which are summated to define the grade. The most representative OA histopathology in the transect taken through the most severe OA lesion was graded in each compartment for the HHGS. In the OACH score there are two components. The grade defines the severity of the disease, in terms of the disease progression of the most severe OA histopathology, and the stage defines the surface extent of this lesion. The OACH grade increases based on the progression of depth involvement of the osteoarthritic lesion from the articular surface to the subchondral bone. Structural surface irregularities are succeeded by increasing loss of cationic stains from superficial to deep articular cartilage layers, followed by increased depth of fissures, erosion of articular cartilage and finally denudation of subchondral bone.

Intra-observer agreement for all observers was graphically represented (Fig. 2). Cone or funnel effects at low or high mean values in either the HHGS or OACH scores were not identified, therefore the relationship between the mean and discrepancy was considered to be consistent and hence the limits of agreement reported were appropriate. The discrepancies between the first and second ratings were larger for the OACH than the HHGS. This could be attributed to the OACH system being new to all raters whereas two raters were expert in the implementation of the HHGS, but also the OACH has an expanded non-continuous scale of 0–24 compared to the HHGS continuous 0–14 scale. Ostergaard *et al.* 1997 assessed the HHGS and stated that they considered the reproducibility to be inadequate, publishing similar reproducibility data in 1999[3,6]. Their method of assessing reproducibility had similarities to Cohen's kappa analysis but their 95% limits of agreement for intra-observer agreement were comparable to those described here[18]. Van Der Sluijs reported HHGS intra-observer agreement as adequate describing similar data to ourselves and Ostergaard[4]. HHGS reproducibility data was published by Custers *et al.* and was described as good as was there description for OACH, this publication provides the only reproducibility data for OACH[5].

Intra-observer and inter-observer reproducibility as defined by the ICC according to Fleiss states <0.40 is poor, 0.40–0.75 is fair to good and ≥0.75 indicates excellent[19]. Therefore our mean HHGS intra-observer ICC was considered excellent and the intra-observer OACH ICC was fair to good. Our inter-observer ICC for the HHGS and OACH was fair to good. These data are supported by Custers *et al.* with HHGS and OACH intra- and inter-observer ICC being described as excellent[5]. Their OACH intra- and inter-observer ICCs were greater

Table V
Intra-observer and inter-observer reproducibility

Intra-observer				Inter-observer			
First & second scores				First score		Second score	
	ICC	CI 95%		ICC	CI 95%	ICC	CI 95%
HHGS							
Observer A	0.816	(0.592, 0.923)	Observer A & B	0.874	(0.710, 0.948)	0.838	(0.636, 0.933)
Observer B	0.905	(0.776, 0.961)	Observer A & C	0.733	(0.439, 0.885)	0.657	(0.314, 0.849)
Observer C	0.687	(0.362, 0.863)	Observer B & C	0.695	(0.374, 0.867)	0.653	(0.307, 0.847)
Mean	0.803±0.110		All observers	0.775	(0.593, 0.895)	0.730	(0.526, 0.872)
OACH							
Observer A	0.542	0.142, 0.789	Observer A&B	0.695	(0.375, 0.867)	0.480	(0.060, 0.756)
Observer B	0.943	0.861, 0.977	Observer A&C	0.387	(−0.056, 0.702)	0.251	(−0.204, 0.617)
Observer C	0.520	0.113, 0.778	Observer B&C	0.296	(−0.157, 0.646)	0.523	(0.117, 0.780)
Mean	0.668 ± 0.238		All observers	0.448	(0.173, 0.701)	0.409	(0.133, 0.674)

CI 95% confidence interval; ± values indicate standard deviation.

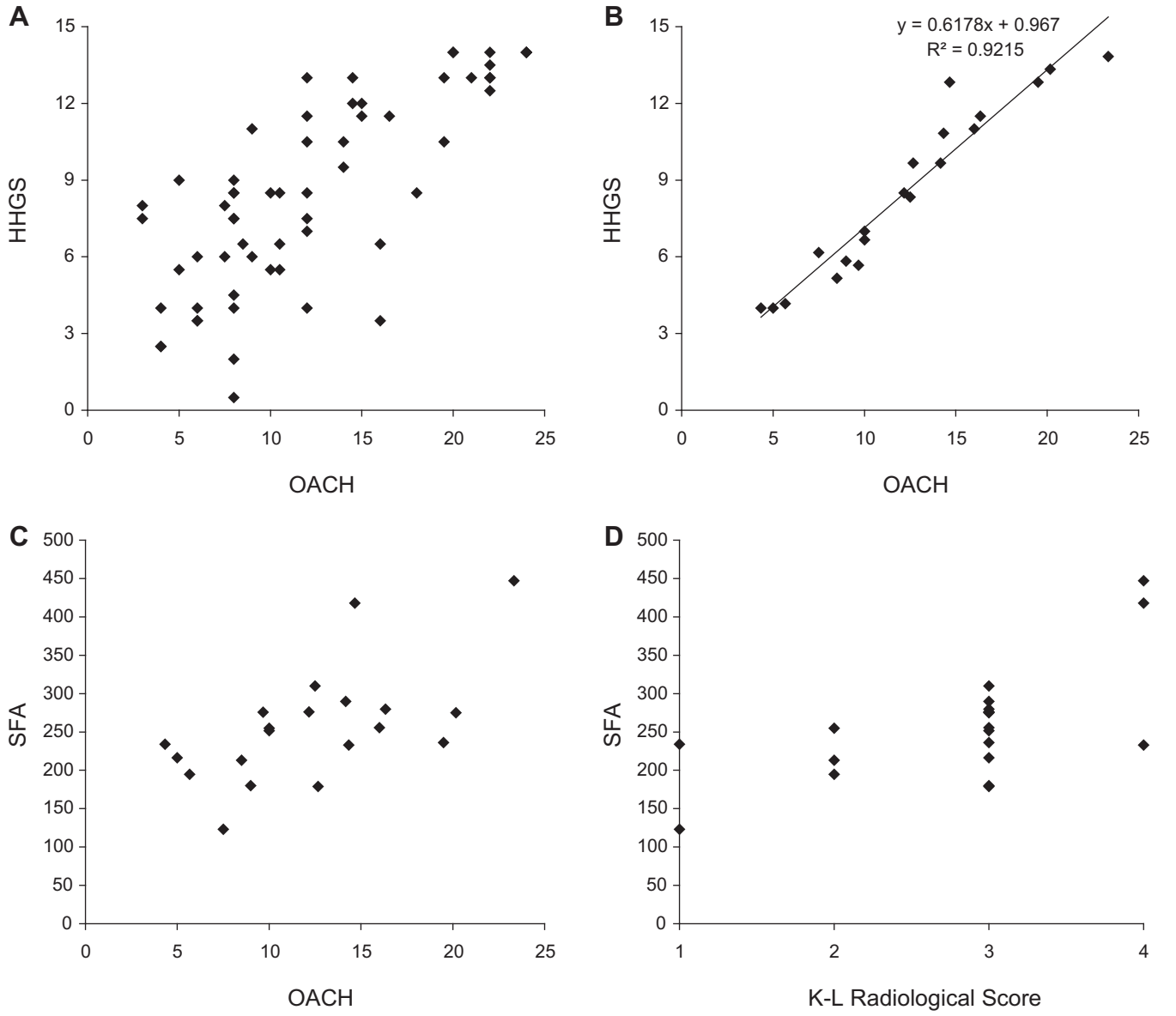

Fig. 3. Validity of HHGS and OACH histopathology. (A) Relationship between HHGS and OACH, all observers (Spearman's ρ 0.757, $P < 0.0001$). (B) Mean HHGS vs mean OACH (Spearman's ρ 0.980, $P < 0.0001$). (C) Validity of OACH histopathology score using SFA score as the reference (Spearman's ρ 0.601, $P = 0.005$). (D) K–L radiological score and SFA chondropathy score correlation (Spearman's ρ 0.537, $P = 0.015$).

than our values which may result from >50% of their specimens originating from the contralateral limb which we expect in their animal model was essentially comprised of normal cartilage; this we consider may result in less scorer variability. Excellent intra-observer and inter-observer ICC was also reported by Acebes *et al.* for HHGS[20].

The reliability of both grading systems was comparable on each scoring occasion for both the OACH and HHGS as identified using a two-way random effect ANOVA, type consistency, average measures ICC. Both OA histopathology systems, on average for the three observers at each scoring occasion had similar reliability. HHGS average measures ICCs of 0.912 and 0.890 on first and second scoring occasions respectively and values of 0.708 and 0.675 for OACH. These reliability statistics are similar to published values, quoted as Cronbach's alpha, for OACH although we obtained marginally higher values for HHGS[5]. Extending these ICCs to an average score from all observers including both scoring occasions increases the reliability estimates to 0.946 and 0.853 for HHGS and OACH.

A highly significant correlation between HHGS and OACH was identified through Spearman's correlation coefficient. This statistic involves ranking, therefore the direct relationship between values was further investigated using linear regression. When mean values for each histopathology grading system were plotted a linear regression with a coefficient of 0.92 was observed.

An accurate description of the reference standard is fundamental to validation. For example the definition of the normal group influences complications such as those arising from the discrimination of degenerative non-osteoarthritic changes and the definition of the osteoarthritic group can address asymptomatic osteoarthritic cartilage[6]. When the definition of the validation group is based upon macroscopically normal articular cartilage vs that with OA pathology the histopathology grading systems have the capacity to discriminate[3]. Validation of histopathology, scoring systems have often used macroscopically graded chondropathy as a reference, and these have included the Beguin and Locker and the SFA score, where sensitivity and specificity were also investigated[20]. Significant correlations have

been reported between the HHGS and chondropathy score, where the specimen transect was consistently made at the midpoint of each tibial plateau[21]. Our OACH histopathology data also correlated with SFA chondropathy scores ($\rho = 0.601$, $P = 0.005$). OACH or HHGS correlation with SFA is unlikely to be highly significant as a single histopathology transect is not likely to be representative of the entire tibial articular surface due to OA lesions tending to be focal. Therefore validating OA histopathology using a visual analogue chondropathy score attributed to the entire compartment may not be the most appropriate validation reference. The correlation between SFA score and radiological score was similar (SFA − K–L $\rho = 0.537$, $P = 0.015$ and SFA − LDA $\rho = 0.573$, $P = 0.008$), therefore indicating that chondropathy is not highly indicative of radiological OA pathology[22,23].

There are advantages and disadvantages to both the HHGS and the OACH[24]. Primary limitations include that the HHGS does not define how a joint surface essentially denuded of articular cartilage should be scored, and the transition between regions of different OA severity can be indistinct and therefore cause difficulty in assigning OACH stage. Also as each OACH grade has multiple associated features in the grading methodology definition, contradictory observations can occur; for example the extent of Safranin-O depletion may not support the depth of fissures defined for Grades 2 and 3 and therefore an informed decision is required. In these cases we scored according to the grade key feature. With respect to the utility of the HHGS and OACH we found that the HHGS took longer to grade. This was primarily because several of the criteria required high magnification examination, in particular this was the examination for breach of the tidemark by blood vessels, whereas all the key features of the OACH assessment can be made at low power.

In conclusion, both the HHGS and OACH systems are reproducible and reliable methods of grading OA histopathology slides, using a three-observer, two-rating, experimental protocol. The linear relationship between HHGS and OACH mean values in conjunction with their Spearman's correlation provides validation of the OACH scoring system for assessment of human tibiofemoral OA pathology.

Contributions

RP and BES conceived and designed the research based on pilot work conducted with KSS. The study material was provided by TK and BES. TK and RP carried out sample preparation and histology. The histopathology scoring was conducted by RP, TK and BES. RP conducted the data analysis and the interpretation was performed in conjunction with BES. The article was drafted by RP and critically reviewed by RP and BES. Funding was obtained by RP and BES. RG Pearson and BE Scammell take responsibility for the integrity of the work as a whole, from inception to finished article.

Conflict of interest

None of the authors have competing interests.

Acknowledgements

The authors would like to thank the Nottingham University Hospital Charity for providing the funds for laboratory consumables (RAP004). We would also like to thank Special Lecturer Michael Seagrave, University of Nottingham, for his kind help with the statistical analysis, particularly during the preparation of the manuscript.

References

1. Roemer FW, Frobell R, Hunter DJ, Crema MD, Fischer W, Bohndorf K, *et al.* MRI-detected subchondral bone marrow signal alterations of the knee joint: terminology, imaging appearance, relevance and radiological differential diagnosis. Osteoarthritis Cartilage 2009;17(9):1115–31.
2. Mankin HJ, Dorfman H, Lippiell L, Zarins A. Biochemical and metabolic abnormalities in articular cartilage from osteoarthritic human hips .2. Correlation of morphology with biochemical and metabolic data. J Bone Joint Surg Am 1971;A 53(3):523–37.
3. Ostergaard K, Andersen CB, Petersen J, Bendtzen K, Salter DM. Validity of histopathological grading of articular cartilage from osteoarthritic knee joints. Ann Rheum Dis 1999;58(4):208–13.
4. Van Der Sluijs JA, Geesink RGT, Van Der Linden AJ, Bulstra SK, Kuyer R, Drukker J. The reliability of the mankin score for osteoarthritis. J Orthop Res 1992;10(1):58–61.
5. Custers RJH, Creemers LB, Verbout AJ, van Rijen MHP, Dhert WJA, Saris DBF. Reliability, reproducibility and variability of the traditional Histologic/Histochemical Grading System vs the new OARSI Osteoarthritis Cartilage Histopathology Assessment System. Osteoarthritis Cartilage 2007;15:1241–8.
6. Ostergaard K, Petersen J, Andersen CB, Bendtzen K, Salter DM. Histologic/histochemical grading system for osteoarthritic articular cartilage – reproducibility and validity. Arthritis Rheum 1997;40(10):1766–71.
7. Pritzker KPH, Gay S, Jimenez SA, Ostergaard K, Pelletier JP, Revell PA, *et al.* Osteoarthritis cartilage histopathology: grading and staging. Osteoarthritis Cartilage 2006;14(1):13–29.
8. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. J Bone Joint Surg Br 1998;80B(1):63–9.
9. Nagaosa Y, Mateus M, Hassan B, Lanyon P, Doherty M. Development of a logically devised line drawing atlas for grading of knee osteoarthritis. Ann Rheum Dis 2000;59(8):587–95.
10. Wilkinson CE, Carr AJ, Doherty M. Does increasing the grades of the knee osteoarthritis line drawing atlas alter its clinimetric properties? Ann Rheum Dis 2005;64(10): 1467–73.
11. Collins DH, McElligott TF. Sulphate ((SO_4)-S-35) uptake by chondrocytes in relation to histological changes in osteoarthritic human articular cartilage. Ann Rheum Dis 1960;19(4):318–30.
12. Brismar BH, Wredmark T, Movin T, Leandersson J, Svensson O. Observer reliability in the arthroscopic classification of osteoarthritis of the knee. J Bone Joint Surg Br 2002;84B(1):42–7.
13. Dougados M, Ayral X, Listrat V, Gueguen A, Bahuaud J, Beaufils P, *et al.* The SFA system for assessing articular-cartilage lesions at arthroscopy of the knee. Arthroscopy 1994;10(1):69–77.
14. Kang Q, LaBreck J, Gruber H, An Y. Histological techniques for decalcified bone and cartilage. In: An Y, Martin K, Eds. Handbook of Histology Methods for Bone and Cartilage. Totowa, NJ: Humana Press Inc; 2003:217–8.
15. Bland JM, Altman DG. Statistical-methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307–10.
16. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1(1):30–46.
17. Bland JM, Altman DG. Cronbach's alpha. Br Med J 1997;314 (7080):572.
18. Brennan P, Silman A. Statistical-methods for assessing observer variability in clinical measures. Br Med J 1992;304 (6840):1491–4.
19. Fleiss J. The Design and Analysis of Clinical Experiments. John Wiley & Sons Inc; 1986. p. 7.
20. Acebes C, Roman-Blas JA, Delgado-Baeza E, Palacios I, Herrero-Beaumont G. Correlation between arthroscopic and histopathological grading systems of articular cartilage lesions in

knee osteoarthritis. Osteoarthritis Cartilage 2009;17(2): 205–12.
21. Walsh DA, Yousef A, McWilliams DF, Hill R, Hargin E, Wilson D. Evaluation of a Photographic Chondropathy Score (PCS) for pathological samples in a study of inflammation in tibiofemoral osteoarthritis. Osteoarthritis Cartilage 2009;17(3): 304–12.
22. Fife RS, Brandt KD, Braunstein EM, Katz BP, Shelbourne KD, Kalasinski LA, *et al*. Relationship between arthroscopic evidence of cartilage damage and radiographic evidence of joint space narrowing in early osteoarthritis of the knee. Arthritis Rheum 1991;34(4):377–82.
23. Kijowski R, Blankenbaker D, Stanton P, Fine J, De Smet A. Arthroscopic validation of radiographic grading scales of osteoarthritis of the tibiofemoral joint. AJR Am J Roentgenol 2006;187 (3):794–9.
24. Rutgers M, van Pelt MJP, Dhert WJA, Creemers LB, Saris DBF. Evaluation of histological scoring systems for tissue-engineered, repaired and osteoarthritic cartilage. Osteoarthritis Cartilage 2009;18(1):12–23.